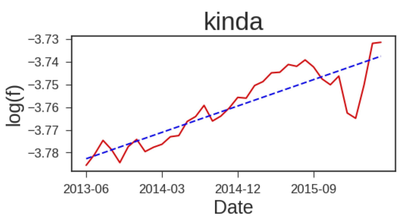
2017

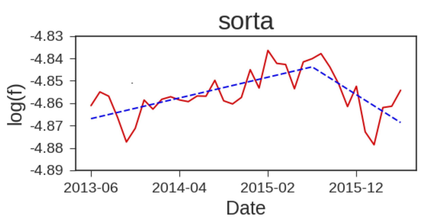
Social and distributional predictors of the success of lexical innovations in online writing

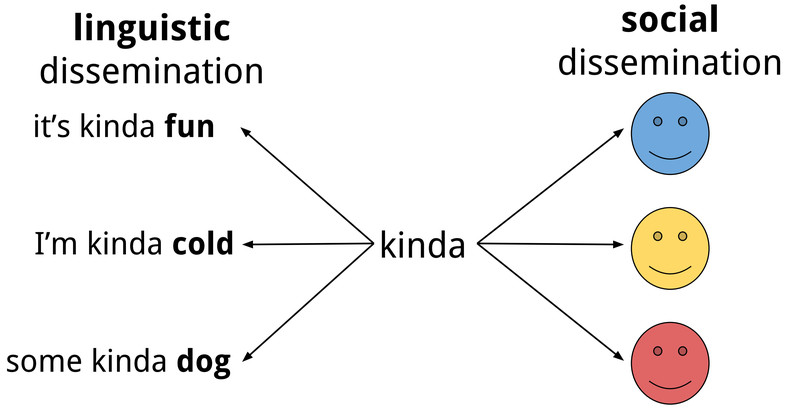
Ian Stewart and Jacob Eisenstein, Georgia Institute of Technology, NWAV 46

What makes innovations succeed?

Online communities are a breeding ground for lexical innovations [3], both successful and failed. The success and failure of innovations is often attributed to the diversity of contexts (**social** and **linguistic**) in which they appear [4]: innovations that "disseminate" across diverse contexts often succeed [1, 5]. This claim has yet to be operationalized and tested on a wide variety of innovations in an online environment.







RQ1: Do successful lexical innovations exhibit higher dissemination across linguistic contexts than failed innovations?
RQ2: Do successful lexical innovations tend to have higher dissemination across social contexts than failed innovations?

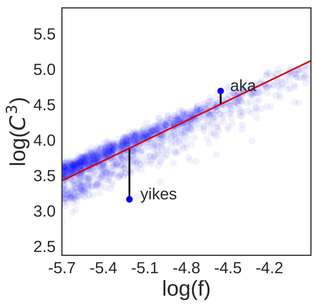
Data and methods

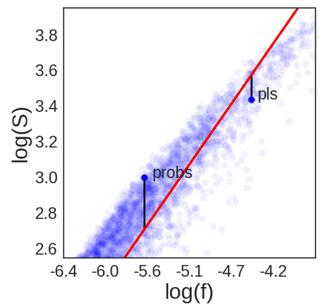
Data	Description
Source	Reddit (English)
Timeframe	2013/06 - 2016/05
Posts	1.6 billion
Tokens	57 billion
Users	14 million
Subreddits	330,000
Threads	100 million
Successful innovations	1451
Failed innovations	600

Innovation type	Examples
acronym	*af, aka, btw*
blend/compound	*fanart, sorta*
clipping	*awk, pls, probs*
derivation	*cringy, ghosting*
onomatopoeia	*aw, boop*
respelling	*boi, succ, tho*
vulgarity/exclamation	*dang, yikes*

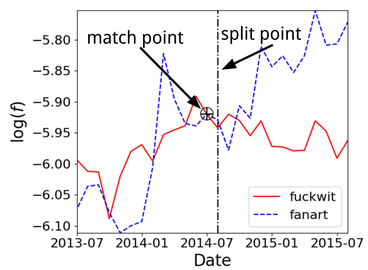
Context dissemination: D^C = ratio of unique trigram contexts (C^3) to expected count.

Social dissemination among users, subreddits, threads [2]: D^U, D^S, D^T = ratio of actual to expected count.





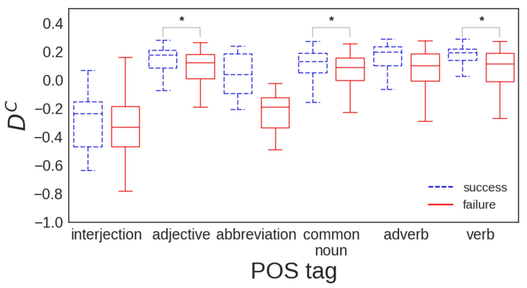
We match successful and failed innovations on frequency before the split point and learn a classifier to predict innovation success.



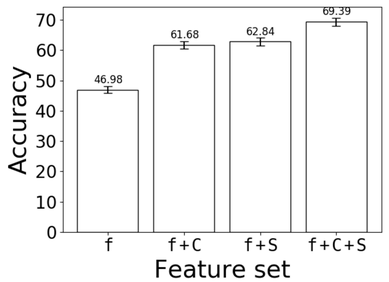
Findings

RQ1: Higher context dissemination predicts innovation success. Consistent across part-of-speech categories (generated with POS tagger trained on Twitter [2]).
RQ2: Higher user and subreddit dissemination predicts innovation success; thread dissemination is insignificant.

Predictor	β
f	0.493 (***)
D^C	3.82 (***)
D^U	3.29 (***)
D^S	1.94 (***)
D^T	-0.769



Context+social dissemination (f+C+S) contributes more to predicting success than context or social alone: social and linguistic factors contribute differently to success.



Our work validates the **Diversity** criterion of the FUDGE model for predicting successful innovations [4]. It also supports the need for comparison tests when studying language change, namely comparing variables at different stages of change. Future work will investigate more nuanced versions of context dissemination (syntactic, semantic contexts).

References

[1] Altmann, E. G., Pierrehumbert, J. B., & Motter, A. E. (2011). Niche as a determinant of word fate in online groups. *PLoS ONE*, *6*(5), 1–12.
[2] K. Gimpel et al. (2011). Part-of-Speech Tagging for Twitter: Annotation, Features, and Experiments. *Proceedings of ACL*, 42–47.
[3] Grieve, J., Nini, A., & Guo, D. (2016). Analyzing lexical emergence in Modern American English online. *English Language and Linguistics*, *20*(2), 1–29.
[4] Metcalf, A. (2004). *Predicting new words: The secrets of their success* (Vol. 18). New York: Houghton Mifflin.
[5] Partington, A. (1993). Corpus evidence of language change: The case of the intensifier. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and Technology: In Honour of John Sinclair* (pp. 177–192). Philadelphia: John Benjamins Publishing.

Full paper (in submission): https://arxiv.org/abs/1709.00345